Overcome Modal Bias in Multi-modal Federated Learning via Balanced Modality Selection

Yunfeng Fan[1], Wenchao Xu[1,*], Haozhao Wang[2], Fushuo Huo,
Jinyu Chen, and Song Guo[3]

[1]PolyU, [2]HUST, [3]HKUST, E-mail: yunfeng.fan@connect.polyu.hk

2024/9/4

Overview

1. Introduction

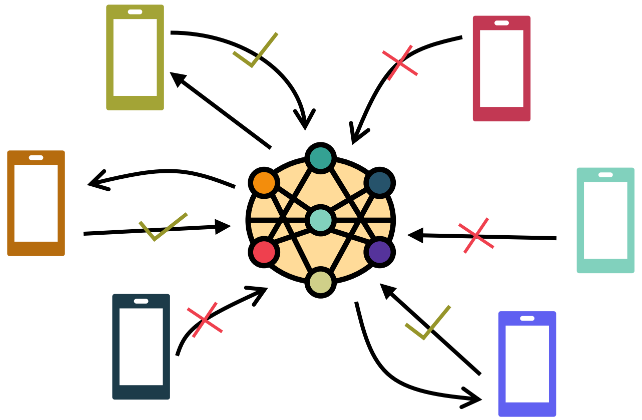
2. Method

3. Results

4. Conclusion

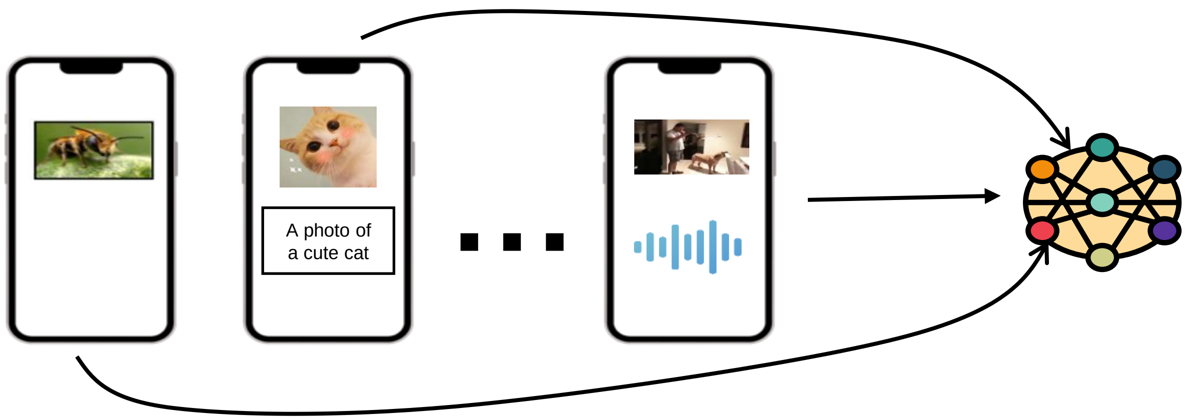
Introduction

Federated Learning (FL)



Collaboratively learn and aggregate knowledge from data that has been collected by, and resides on, a number of remote devices or servers.

Multi-modal Federated Learning (MFL)



Each client contains various types and numbers of modalities of data, making it challenging because of the inter-modal interactions during the MFL training.

Introduction

Observation:

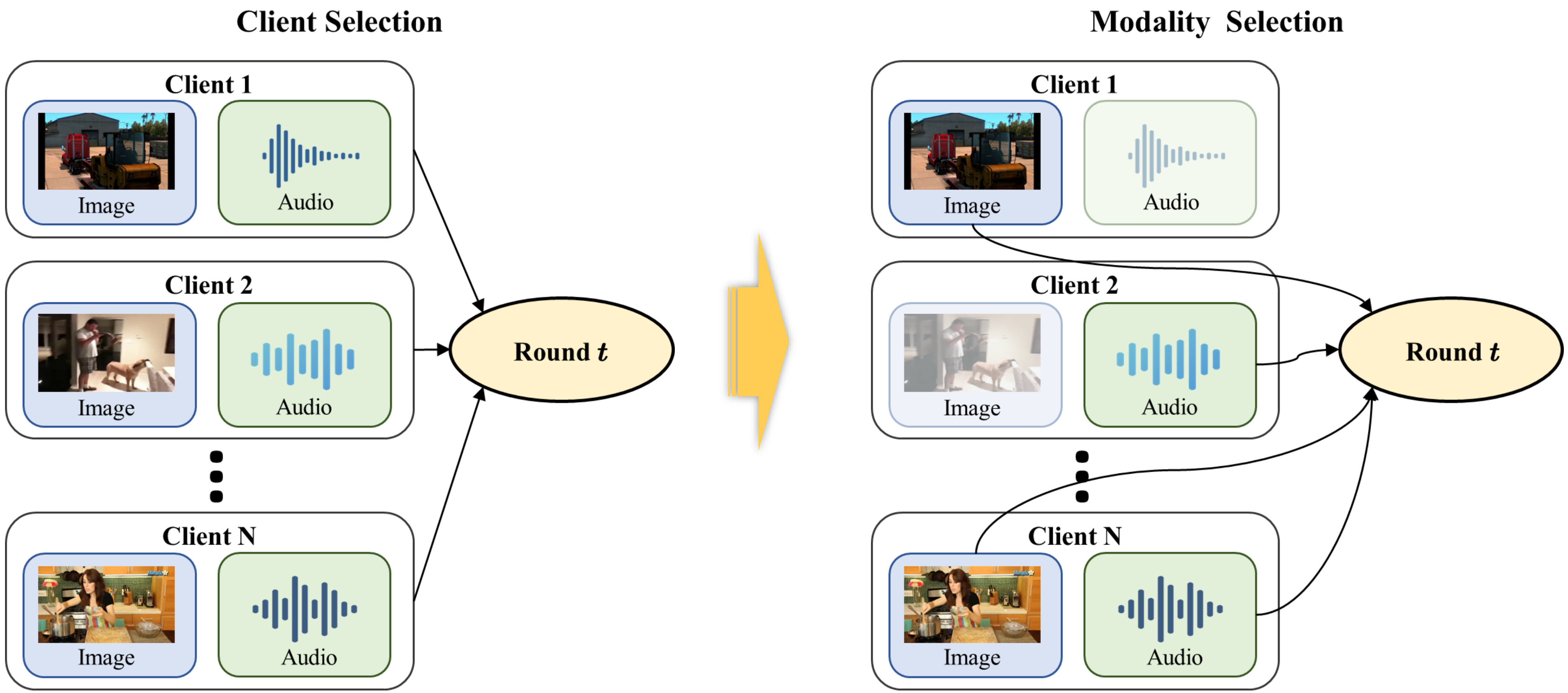
The effectiveness of traditional client selection methods diminishes when dealing with clients with multi-modal data as the inter-modal interactions during MFL training are neglected.

Modality Imbalance

Dataset	CREMA-D [4]			AVE [33]		
Method	A	V	A-V	A	V	A-V
Local	41.9	20.4	39.6	33.4	16.7	35.2
FedAvg	51.2	20.6	50.7	61.1	26.8	62.2
pow-d [6]	51.5	20.4	50.5	61.9	26.9	62.5
DivFL [3]	**52.3**	21.1	51.7	**62.7**	25.3	63.3
FedAvg-0.2	50.6	28.6	52.4	60.6	29.6	63.4
FedAvg-0.5	50.5	34.6	55.7	58.7	30.0	60.7
FedAvg-0.8	48.1	**50.9**	61.2	56.4	31.8	58.5
BMSFed	51.0	41.9	**64.5**	59.7	**40.2**	**64.7**

Introduction

Can we design a new selection scheme in MFL that can overcome the modal bias and exploit each modality comprehensively?



Introduction

Observation:

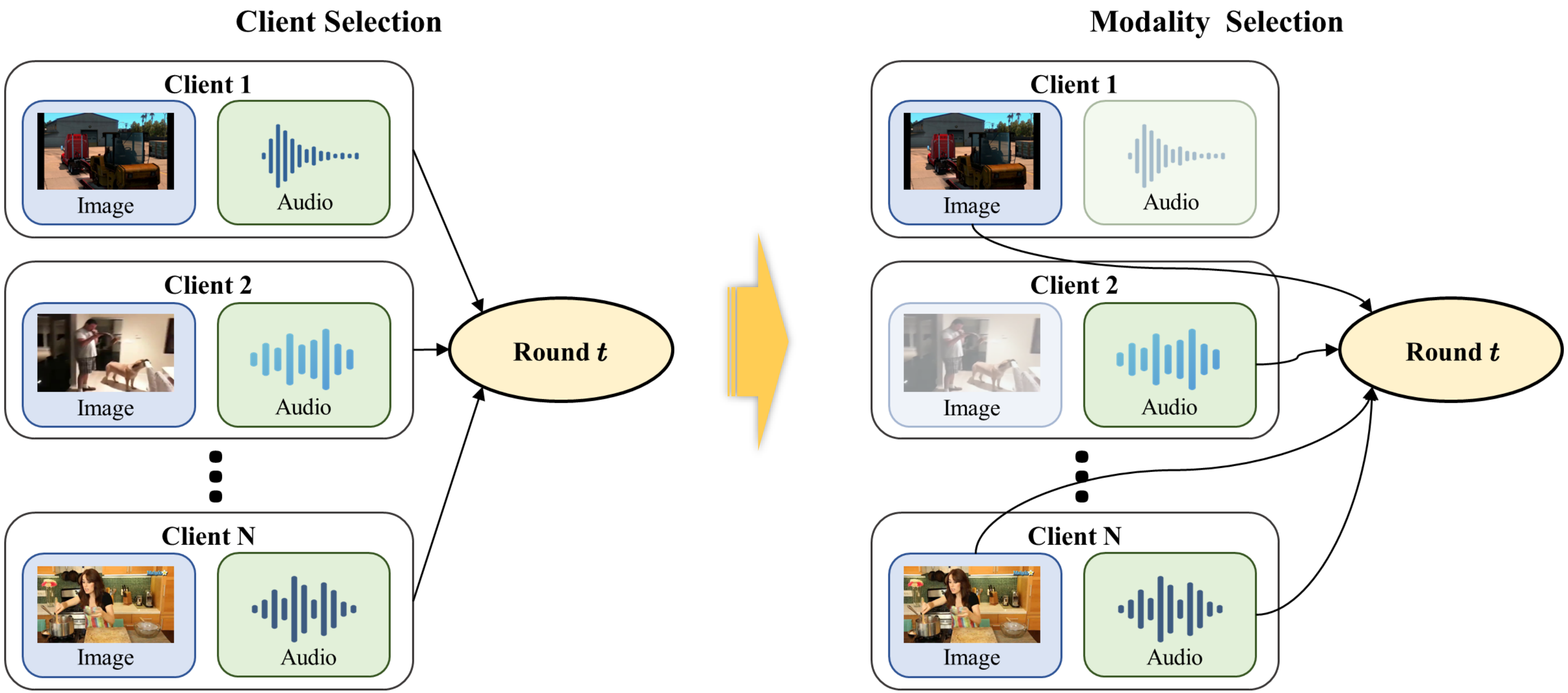
The effectiveness of traditional client selection methods diminishes when dealing with clients with multi-modal data as the inter-modal interactions during MFL training are neglected.

Modality Imbalance

Dataset	CREMA-D [4]			AVE [33]		
Method	A	V	A-V	A	V	A-V
Local	41.9	20.4	39.6	33.4	16.7	35.2
FedAvg	51.2	20.6	50.7	61.1	26.8	62.2
pow-d [6]	51.5	20.4	50.5	61.9	26.9	62.5
DivFL [3]	**52.3**	21.1	51.7	**62.7**	25.3	63.3
FedAvg-0.2	50.6	28.6	52.4	60.6	29.6	63.4
FedAvg-0.5	50.5	34.6	55.7	58.7	30.0	60.7
FedAvg-0.8	48.1	**50.9**	61.2	56.4	31.8	58.5
BMSFed	51.0	41.9	**64.5**	59.7	**40.2**	**64.7**

Introduction

Can we design a new selection scheme in MFL that can overcome the modal bias and exploit each modality comprehensively?



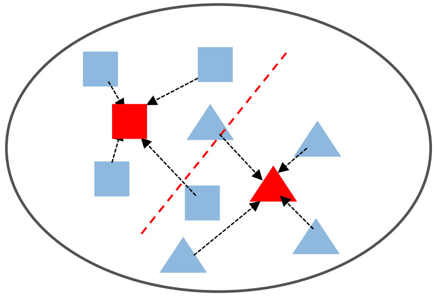
Overview

1. Introduction
2. **Method**
3. Results
4. Conclusion

Method-BMSFed

1. Local Imbalance Alleviation

Leverage class prototypes to calibrate the gradient directions to avoid the inter-modal interference, addressing the inadequate information exploitation on the local side.



Modal Enhancement (ME) Loss

$$\mathcal{L}_{ME}^{k}(v_A) = \mathbb{E}_{(x_i^A, y)\in\mathcal{D}_k} \log\left[\frac{\exp(-d(z_i^A, c_y^{GA}))}{\sum_{j=1}^{Y}\exp(-d(z_i^A, c_j^{GA}))}\right]$$

$$F_k(v_A, v_I) = \begin{cases} \mathcal{L}_{CE}^{k}(v_A, v_I) + \gamma^k \mathcal{L}_{ME}^{k}(v_A) & \rho_I^k \leq 1 \\ \mathcal{L}_{CE}^{k}(v_A, v_I) + \beta^k \mathcal{L}_{ME}^{k}(v_I) & \rho_I^k > 1 \end{cases}$$

where ρ_I^k is the imbalance ratio

Method-BMSFed

2. Balanced Modality Selection

Assume modality I is weak, local loss for multi-modal and uni-modal clients is:

$$\text{multi-modal}: F_k\left(v_A, v_I\right) = \mathcal{L}_{CE}^k\left(v_A, v_I\right) + \beta^k \mathcal{L}_{ME}^k\left(v_I\right)$$

$$\text{uni-modal}: F_k\left(v_A\right) = \mathcal{L}_{CE}^k\left(v_A\right), F_k\left(v_I\right) = \mathcal{L}_{CE}^k\left(v_I\right) + \beta^k \mathcal{L}_{ME}^k\left(v_I\right)$$

Diverse client selection via submodularity:

$$\sum_{k\in[N]} \nabla F_k\left(v_A, v_I\right) = \sum_{k\in[N]} \begin{bmatrix} \nabla F_k\left(v_A, v_I\right) - \nabla F_{\sigma_M(k)}\left(v_A, v_I\right) \\ -\nabla F_{\sigma_A(k)}\left(v_A\right) - \nabla F_{\sigma_I(k)}\left(v_I\right) \end{bmatrix}$$
$$+ \sum_{k\in S_M} \gamma_k^M \nabla F_k\left(v_A, v_I\right) + \sum_{k\in S_A} \gamma_k^A \nabla F_k\left(v_A\right) + \sum_{k\in S_I} \gamma_k^I \nabla F_k\left(v_I\right)$$

where σ_M, σ_A and σ_I map $V \to S_M, S_A, S_I$

$$S_M \cap S_A = S_A \cap S_I = S_M \cap S_I = \emptyset.$$

Method-BMSFed

Since modality I is weak here, we omit the uni-A clients as the multi-modal gradient is dominated by modality A

$$\sum_{k\in[N]} \min_{i\in S_M, j\in S_I} \left\|\nabla F_k(v_A, v_I) - \gamma_i^M \nabla F_i(v_A, v_I) - \gamma_j^I \nabla F_j(v_I)\right\|$$

$$= \sum_{k\in[N]} \min_{i\in S_M, j\in S_I} \left\| \begin{array}{c} \nabla \mathcal{L}_{CE}^k(v_A, v_I) + \nabla \beta^k \mathcal{L}_{ME}^k(v_I) - \nabla \mathcal{L}_{CE}^i(v_A, v_I) \\ -\nabla \beta^i \mathcal{L}_{ME}^i(v_I) - \nabla \mathcal{L}_{CE}^j(v_I) - \nabla \beta^j \mathcal{L}_{ME}^j(v_I) \end{array} \right\|$$

Gradient decoupling

$$\leqslant \sum_{k\in[N]} \min_{i\in S_M} \left\|\nabla \mathcal{L}_{CE}^k(v_A, v_I) - \nabla \mathcal{L}_{CE}^i(v_A, v_I)\right\|$$

$$+ \sum_{k\in[N]} \min_{i\in S_M, j\in S_I} \left\| \begin{array}{c} \nabla \beta^k \mathcal{L}_{ME}^k(v_I) - \nabla \beta^i \mathcal{L}_{ME}^i(v_I) \\ -\nabla \mathcal{L}_{CE}^j(v_I) - \nabla \beta^j \mathcal{L}_{ME}^j(v_I) \end{array} \right\|$$

$$\triangleq G(S_M) + G(S_M \cup S_I)$$

Method-BMSFed

Solve the two submodular functions with the stochastic greedy algorithm

- The type of selected client according to $G(S_M \cup S_I)$ should be specified;
- The separated selection strategy pays less attention to the global modal bias

Conflict Resolution Strategy

$$S_M \leftarrow S_M \cup k_1^*, k_1^* \in \underset{k \in \mathrm{rand}(V \backslash S_M \backslash S_I, \mathrm{s})}{\arg\max} \left[\bar{G}(S_M) - \bar{G}(\{k\} \cup S_M)\right]$$

$$\begin{cases} if\ k_1^* = k_2^*, S_M \cup k_2^*; \\ if\ k_1^* \neq k_2^*, \begin{cases} S_I \cup k_2^*, if\ \rho_I^k > \chi \\ S_M \cup k_2^*, if\ \rho_I^k \leqslant \chi \end{cases} \end{cases}$$

$$k_2^* \in \underset{k \in \mathrm{rand}(V \backslash S_M \backslash S_I, \mathrm{s})}{\arg\max} \left[\bar{G}(S_M \cup S_I) - \bar{G}(\{k\} \cup S_M \cup S_I)\right]$$

Overview

1. Introduction

2. Method

3. Results

4. Conclusion

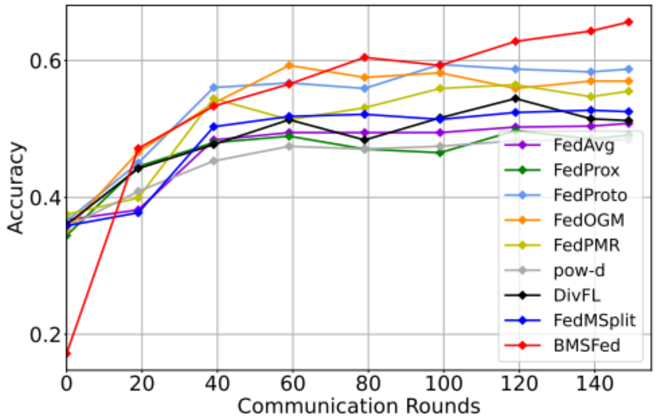
Results

1. Performance Comparison to SOTA Baselines

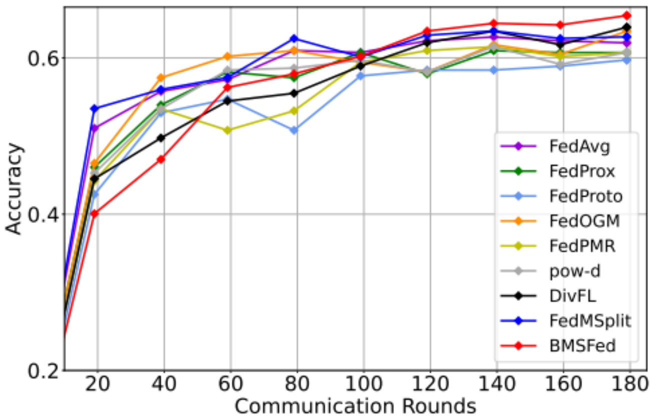
Dataset	CREMA-D		AVE		CG-MNIST		ModelNet40	
Method	IID	non-IID	IID	non-IID	IID	non-IID	IID	non-IID
FedAvg	50.7	49.8	62.2	59.7	42.3	41.7	87.2	86.5
FedProx	51.0	49.0	62.6	59.9	42.9	43.6	86.9	87.1
FedProto	58.7	54.0	61.7	58.8	51.5	51.4	87.5	87.2
FedOGM	56.9	56.4	62.8	59.3	57.2	53.0	87.6	87.0
FedPMR	55.5	55.1	63.1	61.6	66.1	63.3	87.6	**87.7**
pow-d	50.5	50.7	62.5	60.0	41.2	40.3	86.8	86.2
DivFL	51.7	50.8	63.3	59.6	43.0	42.1	86.5	86.4
FedMSplit	52.4	51.6	62.4	60.8	43.5	50.9	87.5	87.4
BMSFed	**64.5**	**61.6**	**64.7**	**62.1**	**70.2**	**66.7**	**88.7**	87.5

2. Uni-modal Performance Comparison

Dataset	CREMA-D				AVE			
Setting	IID		non-IID		IID		non-IID	
Method	A	V	A	V	A	V	A	V
FedAvg	51.2	20.6	50.7	20.2	61.1	26.8	61.4	26.4
FedProx	51.3	20.2	50.1	22.0	60.4	27.1	61.2	26.9
FedProto	50.2	35.3	48.6	39.1	55.7	36.8	59.7	32.8
FedOGM	50.5	35.7	48.8	30.2	58.7	28.8	59.4	29.4
FedPMR	51.5	38.7	50.1	35.9	61.7	39.6	61.7	35.3
pow-d	51.5	20.4	51.6	18.8	61.9	26.9	60.1	27.1
DivFL	**52.3**	21.1	**52.1**	22.7	**62.7**	25.3	61.6	26.3
FedMSplit	52.0	21.8	50.8	21.6	61.3	26.9	**62.3**	28.7
BMSFed	51.0	**41.9**	49.3	**41.4**	59.7	**40.2**	60.2	**38.6**



(a) CREMA-D under IID



(b) AVE under IID

Thanks for your attention!